DOCUMENT RESUME

ED 112 900 IR 002 611

AUTHOR Regazzi, John J.
TITLE The Development of Some Automated Techniques of Information Retrieval in Family Planning and Population Libraries.
PUB DATE Jun 75
NOTE 19p.; Paper presented at the Medical Library Association Annual Meeting (74th, Cleveland, Ohio, May 30 through June 5, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Computer Programs; Data Bases; Family Planning; *Information Retrieval; *Information Systems; Librarians; *Library Automation; On Line Systems; Population Trends; *Special Libraries; Speeches; Thesauri
IDENTIFIERS Carolina Population Center; *POPINFORM; Population Information

ABSTRACT

The Carolina Population Institute off-line system which searches a data base drawn from its library holdings and POPINFORM--an on-line system containing a number of data bases from both university and government organizations--are described. The development of these automated information retrieval systems from a disjoint manual system among a variety of institutions to a singular data base is discussed. The preparation and planning techniques required for any library to convert its manual routines to automated ones are outlined. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED112900

THE DEVELOPMENT OF SOME AUTOMATED TECHNIQUES OF INFORMATION RETRIEVAL IN FAMILY PLANNING AND POPULATION LIBRARIES

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

by

John J. Regazzi*
Systems Librarian
Northern Illinois University

*Formerly: Library Director
Association for Voluntary Sterilization

IR 002 611

ABSTRACT

Two automated information retrieval systems for population and family planning are described. (1) The Carolina Population Institute has an off-line system which searches a data base drawn from its library holdings; (2) POPINFORM is an on-line system containing a number of data bases from both university and government organizations. The development of these automated information/retrieval systems from a disjoint manual system among a variety of institutions to a singular data base is discussed in detail. This preparation and planning required for any library to convert its manual routines to automated ones is therefore also treated.

Introduction

As is common in other specialized areas, automated information retrieval is often brought about because of the need for access to very specific information, relatively quickly. There are presently two well developed computer-based information systems in the field of population and family planning. The design of each is quite different, yet both contain bibliographic data which can be retrieved in a variety of ways. In order to understand better the development of these automated systems, a brief discussion of their evaluation has been included.

Automated Information Retrieval Systems for Population/Family Planning

System	Sponsoring Institutions	Characteristics
1. Technical Information service, CPC	a. Carolina Population Center, University of North Carolina	a. Offline system b. Batch searching c. Author, title, and subject searches d. Bibliographic citations without abstracts
2. POPINFORM	a. Center for Disease Control b. East West Center c. George Washington University d. Columbia University	a. Online system b. Online searching c. Author, title, and subject searches d. Bibliographic citations available with abstracts e. Online and offline listings available

Illustration 1

Background

The fields of populations and family planning, owing in part to the recent infusions of research efforts, program development, and money fit the need to develop a systematic information system. Though this has been done in a variety of ways by different countries and special interest groups within the field, every information center, regardless of scope, had to deal with similar problems of acquisitions and storage of materials, and inputting and retrieving data. Moreover, the development of a thesaurus is critical to any information service. Though a variety of thesauri have been developed along special interest lines, the one produced by Columbia University's International Institute for the Study of Human Reproduction is most comprehensive and illustrates how good planning in the original design of a tool, intended to supplement a manual operation can later be utilized with an automated system.

Thesaurus Development

The Institute's thesaurus, entitled Fertility Modification Thesaurus with Focus on Evaluation of Family Planning Programs,[1] differs from the National Library of Medicine's Medical Subject Headings (MESH) in content and scope, but it is similar in format and certain terminology. The Association of Voluntary Sterilization, along with other family planning libraries, utilized this thesaurus as the basis of its own.

The arrangement of the Fertility Modification Thesaurus is in four main parts: (1) the thesaurus proper;
(2) a list of discriptors only;
(3) a list of categories; and
(4) a hierarchy list.

The first section, an alphabetical arrangement of terms, forms the main portion of the thesaurus. For each descriptor, or term, the entry includes a definition or scope note where desirable or necessary (this appears in italics below the entry). A "Broader Term" (BT) denotes the term immediately above the description in its hierarchy, and "Narrower Term" (NT) signals the term or terms immediately below it in the hierarchy. "Related Term" (RT) signifies a related concept, but one that does not fit into the same hierarchy as that particular description. All terms, however, within that particular hierarchy are related terms, but do not possess RT designations. There are other relational codes. "USE" is a signal to a preferred term from a disallowed term. "USED FOR" is a reference under the preferred term to what the disallowed terminology is. "AND" (when in all caps) signals the use of two preferred terms to express a single disallowed term. In order to illustrate the relationship to MeSH, one or more alphanumeric codes appear along side this description. Those in parenthesis are MeSH designation; however, those not in parenthesis are in the MeSH format, but were compiled for this list. Some terms have both one or more MeSH codes and other codes developed by the Institute. (N5, for example, is a specially developed category for this thesaurus "Family Planning Program Development and Evaluation.").

The second section is a list of preferred terms arranged alphabetically. This can be used as a quick reference for indexers, and if the term is not found here the first list should be consulted.

A third section represents a tree display of each category and preferred term in a hierarchical arrangement. This illustrates where each term falls in the indexing hierarchy; thus insuring the desired level of specificity has been assigned.

The final section included is a simple list of all preferred terms. These are arranged in a single alphabet without regard to word relationships. This section is used as a reference section, or authority list, for the indexers. The comprehensive nature of the thesaurus was such that other family planning or population libraries could use it as their own authority list with little or no adaptation. This was done for example, by the Association of Voluntary Sterilization which merely expanded appropriate sections of the thesaurus.

Manual Applications

At Columbia University and the Association of Voluntary Sterilization, this thesaurus served as the basis of a relatively standardized manual indexing system which could be later converted to a computer based system if so desired. What follows is a brief analysis of the procedures for indexing used at the Association. This discussion is useful as it illustrates a modus operandi which forms the basis of a computer application, and conversely demonstrates how some elementary principles of computer searching can be applied manually.

Materials, including monographs, which are of permanent value to the collection are given full cataloging, i.e., author, title and all added entries. The bulk of these materials are given a unique number and filed according to that number. They were physically filed by the unique number and thus are not classified by subject. Monographs are shelved and arranged by the author's name; cross-references, however, are made from the main file to this shelf position.

Before final processing, the indexer assigns subject terms to the document as necessary. There is no limit to the number which can be assigned to the document, the average is about 7-1/2 terms per title. A subject card, similar to the old uniterm card, is made for each title. One card is made for each allowable subject term, not every term, however. This differentiates it from a uniterm index.

The allowable terms are uniform and singular in concept, though not always in terminology. For example, "PUBLIC ASSISTANCE" is an allowable subject term, the system need not coordinate the terms "Public" and "Assistance" to retrieve document treating this concept. Pragmatic considerations of time and space dictate such pre-coordination, yet it is not without theoretical basis either. A composite grouping of words often refer to a single concept. The composite, as it is represented in the literature, often refers to this concept in a superior way than the coordination of its several terms would. Take for example in the terms "Public Assistance" and "Private Companies". If these terms were coordinated through each of its individual members, it could also include a variety of "False drops" such as a "Private Assistance and Public Companies". Pre-coordination

keeps such retrievals to a minimum.

The unique number of the document is then entered on the individual subject term cards, thus permitting later coordination for retrieval. If an individual, for example, desire material on the "Failure Rate of Vasectomy in India," the corresponding cards, i.e., "Vasectomy", "Failure Rates," and "India" are pulled and those numbers which match are "hits" or represent the desired documents which treat that topic. (See Illustration 2).

The step from this system to a computer-based system was not a big one, but it did require conversion of records to a machine readable form, software to do the searching, and money to finance the project.

Uniterm Procedures for a Manual Indexing System

Vasectomy

0	1	2	3	4	5	6	7	8	9
1000	1001	1002	1003			1006			1119

Failure Rates

0	1	2	3	4	5	6	7	8	9
1000	1001		1003						1119

India

0	1	2	3	4	5	6	7	8	9
	1001								1119

Illustration 2

Carolina Population Center

The Carolina Population Center which holds one of the largest collections of family planning and population literature in the world has effectively converted its holdings to machine-readable form for batch searching. The Center serves both its own institution as well as smaller family planning libraries scattered around the world. The Technical Information Service long ago developed a method of manual searching CPC's holdings for a variety of subjects as well as author, title and added entries. This was accomplished by the use of a coordinate indexing scheme for both journal articles and monographs. Patrons could request searches for a variety of topics with varying levels of specificity and use the system equally.

It was soon realized that these searches if automated could be faster and more economical of staff time. Files were gradually converted to machine readable form, and offline computer searches are available at a nominal cost to the user from the Center. These searches are available to libraries from less developed countries at not cost. Carolina Population Center keeps all of their searches previously done, updating these periodically or upon request. CPC's searches tend to be excellent for retrospective searches, since most of CPC's collection is available in this file. POPINFORM, by comparison, is on-line and composed of several institutions each inputting complimentary data to the data base.

POPINFORM

POPINFORM is an acronym for POPulation INFORMation on-line, and it constitutes a sizeable computer data base for the fields of population and family planning. POPINFORM is inter-disciplinary, as the fields of population and family planning encompass many varied disciplines such as clinical and preventive medicine, demography, biology, economics, and other social sciences to name a few.

Online System Development

The design of POPINFORM is similar to Medline in that the user can access the data base through normal voice grade telephone lines with a variety of terminals.

To use POPINFORM the subscriber uses conventional telephone communications to connect with the computer containing the data base in Washington, D.C. These telephone connections extend locally, nationally, as well as internationally through telecommunications and satellite connections. Manila, for example, has been connected to POPINFORM through such means. The computer prompts: "Welcome to POPINFORM" and "File P is ready for your search," and the user can begin his search employing subject terms which correspond to the thesaurus described above and MeSH terms. The searcher can use boolean logic to attain greater specificity of retrieval. The user can then request specific subfile, if desired. Author, titles and added entries may also be searched through the use of full author or title/author search keys. The citations and abstracts can then be printed on-line

at the terminal, or offline in Washington, D.C. and be sent through the mail. A simple sign-off procedure terminates the session.

System Design

POPINFORM contains population information on a national and international scope first assembled by the Population Information Program of George Washington University, Department of Medical and Public Affairs under a contract from the United States Agency for International Development, and has now been made commercially available through Informatics, Inc. as an on-line interactive real time system for contracted subscribers.

Given the conflate nature of population information, POPINFORM needs to acquire its literature from a variety of sources. Newsletters, newspapers, United Nations documents, unpublished papers and reports, and census data are just some of the kinds of materials which are used in addition to standard bibliographic sources such as periodicals and monographic literature. POPINFORM, if it was to function effectively, needed to devise a way of adequately covering the fields of population and family planning. To attempt to determine a discrete population of family planning/population bibliographic sources was virtually impossible. Consequently, the indexing of a specified list of journals even if done comprehensively would probably not suffice in covering the literature.

Since the library for the Population Information Program (PIP) was already indexing all acquisitions for in-house use, they could transfer their collection into machine readable form and use it as the basis for a computerized data base. PIP was able to combine similar efforts with several other institutions to form the core of POPINFORM.

Scope

The five basic files of POPINFORM are for the most part mutually exclusive. They may be more appropriately thought of as subfiles. They are as follows:

File D (1) Center for Disease Control, Family Planning Evaluation Division (CDC-FPED).

File E (2) East-West Center, East-West Communication Institute (EWCI)

File J (3) George Washington University, Population Information Program (GWU-PIP)

File K (4) George Washington University, Prostaglandin Information Center (GWU-PIC)

File L (5) Columbia University, International Institute for the Study of Human Reproduction.

These files are concurrently searchable through File P - POPINFORM. File P is the designated default file for the system. A user may, if so desired, also only search one of the subfiles exclusively. Presently POPINFORM consists of over 25,000 citations and abstracts, and is growing at the rate of about 1,000 entries per month. POPINFORM has been commercially available for less then one year. (See Illustration 3).

Center for Disease Control, Family Planning Evaluation Division.

CDC-FPED consists of over 6,000 entries with abstracts. This file's indexing terms can be accessed and displayed through the central file. Included in this subfile is U.S. information on family planning evaluation, pregnancy termination, and complications associated with oral contraceptives and IUD's. Entries include journal articles, conference papers, and unpublished reports.

East-West Center, East-West Communication Institute.

EWCI contains some 1,000 entries representing a strong collection of literature on information, education, and communication materials for family planning and population programs. Sources are varied, but include many unpublished reports and documents which are not cited in standard bibliographic sources. Abstracts are not available yet for many of these citizens.

Subfile Profile of POPINFORM

Subfile	Approximate Size	Scope
Center for Disease Control, Family Planning Evaluation Division (CDC-FPED)	6,000	U.S. information of family planning evaluation, pregnancy termination and contraception
East-West Center, East-West Communication Institute (EWCI)	1,000+	Worldwide information on family planning information, education and communication programs.
George Washington University, Population Information Program (GWU-PIP)	9,000	Population and law literature; experimental contraceptive technology
George Washington University, Prostaglandin Information Center	2,500	Literature on prostaglandins relating to fertility control.
Columbia University, International Institute for the Study of Human Reproduction (CUIISHR)	8,000	Family planning program evaluation literature

Illustration 3

George Washington University, *Population Information Program.*

GWU-PIP is the largest of these subfiles and the broadest in subject content. PIP consists of nearly 9,000 entries with abstracts. It includes such topics as experimental and "state of the art" contraceptive technology reports, family planning programs and population law and policy. PIP thesaurus which is based on the National Library of Medicine's MeSH is only available through the subfile. Thus certain secondary terms which are unique to this thesaurus are not available in the conglomerate File P. They can be searched, however, by requesting this subfile itself.

George Washington University, Prostaglandin Information Center.

GWU-PG includes some 2,000 entries with abstracts available directly through File P. This file is the most specialized of the group. The thesaurus as in the PIP file is not available through File P, but can be accessed through the individual subfile. PG consists exclusively of literature on prostaglandins relating to fertility regulation.

Columbia University, International Institute for the Study of Human Reproduction.

CUIISHR is a large bibliographic file of some 8,000 entries. It specializes, however, in the literature of family planning program evaluation. It includes a variety of sources, such as journal articles, unpublished reports and documents. In lieu of an abstract, these citations contain numerous indexing terms from which the user can estimate the contents of the entries.

Costs and Users

Considering the fact that POPINFORM is relatively new, a variety of users around this country and the world are subscribing. Though at present they have less than 75 subscribers, these represent a variety of groups of governmental agencies and private associations. These users comprise various disciplines which not only include population and family planning associations but also public health groups and various health agencies among others.

Presently the basic costs for the system are $50 per hour for online time which is prorated so that you pay for only online time. There is no minimum requirement for online useage. Offline printing of citations and abstracts at the computer site is 20¢ per page, with a $5.00 minimum order.

Users must make arrangements for the purchase or lease of their terminal. POPINFORM is compatible with a variety of teletype, CRT, or IBM 2741 type terminals. If users require the use of POPINFORM's tymshare system which is available throughout the continental United States, there is an additional $10 per hour surcharge prorated as given in the conditions above.

Conclusion

In conclusion, POPINFORM offers a truly centralized information system for a specialized field of literature. Libraries and information centers related to population and family planning should consider its utility. As well, appropriate institutions, no matter how small, should consider the conversion of their own specialized files to machine readable form for inclusion in POPINFORM. One of the strengths of this system is in its subfile design, giving it a great flexibility not always characteristic of automated library systems.

Finally, though there is some duplication between POPINFORM and MEDLINE in terms of the specific citations and abstracts, clearly there is little or no excessive replication of indexing terminology. This replication seems justified and needed due to the wide difference of scope and functions of the two systems.

References

1. Speert, Kathryn H. and Samuel M. Wishik. Fertility Modification Thesaurus with Focus on Evaluation of Family Planning Programs, New York, Columbia University, International Institute for the Study of Human Reproduction, 1973.